SPEAKER RECOGNITION USING LOCAL MODELS

ABSTRACT OF THE DISCLOSURE

[0048]   A system and method for voice recognition is disclosed. The system enrolls speakers using an enrollment voice samples and identification information. An extraction module characterizes enrollment voice samples with high-dimensional feature vectors or speaker data points. A data structuring module organizes data points into a high-dimensional data structure, such as a kd-tree, in which similarity between data points dictates a distance, such as a Euclidean distance, a Minkowski distance, or a Manhattan distance. The system recognizes a speaker using an unidentified voice sample. A data querying module searches the data structure to generate a subset of approximate nearest neighbors based on an extracted high-dimensional feature vector. A data modeling module uses Parzen windows to estimate a probability density function representing how closely characteristics of the unidentified speaker match enrolled speakers, in real-time, without extensive training data or parametric assumptions about data distribution. A smoothing parameter controls the relative contributions of close and far speaker data points to the estimated density.